

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 16:16:39 ; Search time 91.75 Seconds
(without alignments)
24.287 Million cell updates/sec

Title: US-09-331-631A-3_COPY_117_185
Perfect score: 384
Sequence: 1 NRORPQOQYEQCEQRCORH.....EEQREDEKYEERKEEDN 69

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	149.5	38.9	588	VCLB_GOSHI	P09801 gossypium h
2	133	34.6	605	VCLB_GOSHI	P09799 gossypium h
3	113	29.4	1898	TRHY_HUMAN	Q07283 homo sapien
4	110	28.6	1407	TRHY_RABIT	P37709 oryctolagus
5	98	25.5	482	U2R2_HUMAN	Q15696 homo sapien
6	98	25.5	648	KAPC_DICDI	P34099 dictyosteli
7	97	25.3	2124	Y192_HUMAN	Q93074 homo sapien
8	96	25.0	429	APPA_MACPA	P33621 macaca fasc
9	96	25.0	524	SBP_SORBN	Q04672 glycine max
10	95.5	24.9	538	CK11_YEAST	P23291 saccharomyc
11	94.5	24.6	678	GARP_PLAIF	P13816 plasmodium
12	94.5	24.6	877	INCE_CHICK	P53352 gallus gall
13	94.5	24.6	1505	SIMA_DROME	Q24167 drosophila
14	94	24.5	1023	GUF_DROME	P33338 drosophila
15	94	24.5	1085	YAFB_SCHPO	Q09863 schizosacch
16	94	24.5	1349	TRHY_SHEEP	P22793 ovis aries
17	93	24.2	1403	PRO_DROME	P22617 drosophila
18	92.5	24.1	758	YMS3_YEAST	Q03825 saccharomyc
19	92	24.0	1157	Y182_HUMAN	Q14680 homo sapien
20	91	23.7	905	SNF5_YEAST	P18480 saccharomyc
21	90.5	23.6	550	BLSA_HUMAN	Q02832 homo sapien
22	90.5	23.6	573	XET_HUMAN	P15590 zea mays (m
23	90.5	23.6	695	GLB1_MALZE	Q02040 homo sapien
24	90.5	23.6	1048	SR44_RAT	Q05627 rattus norv
25	90.5	23.6	1905	TAGB_DICDI	P54683 dictyosteli
26	90	23.4	338	LEGB_PEA	P14594 pisum sativ
27	89.5	23.3	743	ABRA_PLAIF	P22620 plasmodium
28	88	22.9	1240	YND1_YEAST	P53935 saccharomyc
29	87.5	22.8	708	GBF_DICDI	P36417 dictyosteli
30	87	22.7	1023	CLOC_DROME	Q61735 drosophila
31	86.5	22.5	2038	FSH_DROME	P13709 drosophila
32	86	22.4	386	ARPA_STRPY	P13050 streptococc
33	85.5	22.3	47	AGRP_L0FCY	P56568 lufia cyllin

34	85.5	22.3	471	1	RU17_XENLA	P09406 xenopus lae
35	85.5	22.3	572	1	MOES_LYIVA	P52962 lytechinus
36	85	22.1	285	1	INVO_CANFA	P18174 canis fam11
37	85	22.1	1178	1	MNN4_YEAST	P36044 saccharomyc
38	85	22.1	1344	1	IF3A_MOUSE	P23116 mus musculu
39	85	22.1	1382	1	IF3A_HUMAN	Q14152 homo sapien
40	84.5	22.0	321	1	ABRA_PLAIF	Q23746 plasmodium
41	84	21.9	479	1	U2R1_HUMAN	Q15695 homo sapien
42	83.5	21.7	344	1	INVO_PIG	P18175 sus scrofa
43	83.5	21.7	444	1	CER_DROME	Q02637 drosophila
44	83.5	21.7	605	1	GLCA_SORBN	P13916 glycine max
45	83	21.6	389	1	M49_STRPY	P16947 streptococc

ALIGNMENTS

RESULT 1	ID	VCLB_GOSHI	STANDARD	PRT	588 AA.
AC	P09801				
DT	01-MAR-1989 (rel. 10, Created)				
DT	01-MAR-1989 (rel. 10, Last sequence update)				
DT	15-JUL-1999 (rel. 38, Last annotation update)				
DE	VICILIN C72 PRECURSOR (ALPHA-GLOBULIN B).				
OS	Gossypium hirsutum (Upland cotton).				
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.				
RN	(1)				
RP	SEQUENCE FROM N.A.				
RA	Chlan G.A., Pyle J.B., Legocki A.B., Dure J., III;				
RT	"Developmental biochemistry of cottonseed embryogenesis and germination. XVIII. cDNA and amino acid sequences of the members of the storage protein families."				
RT	Plant Mol. Biol. 7:475-489(1986).				
RL	Plant Mol. Biol. 7:475-489(1986).				
CC	- FUNCTION: SEED STORAGE PROTEIN.				
CC	- SUBCELLULAR LOCATION: CYTOLEDONARY MEMBRANE-BOUND VACUOLAR PROTEIN BODIES.				
CC	- SIMILARITY: TO OTHER 75 SEED STORAGE PROTEINS (PHASEOLIN, VICILIN, CONVICTILIN, CONGLICININ, ETC.).				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL, M16891; AAA33071.1; ..				
DR	PIR, A30838; FMCNAB.				
DR	HSSP, P50477; ICAX.				
DR	INTERPRO, IPR001113; ..				
DR	PFAM, PF00546; Seedstore_7s; 1.				
FT	Seed storage protein: Signal.				
FT	SIGNAL	1	25		
FT	CHAIN	1	588		
FT	SEQUENCE	588 AA; 69729 MW; 63E69B29A8DADEB CRC64;			

Query Match 38.9%; Score 149.5; DB 1; Length 588;
Best Local Similarity 33.7%; Pred. No. 1.3e-05;
Matches 30; Conservative 21; Mismatches 15; Indels 23; Gaps 3;

QY	1	NRORPQOQYEQCEQRCORHETPRHMTQRCGRYRKERKQQRKYEDQ-----	53
DB	78	HRPEDPQRREYECQOBCR--QOEERORPQOCORCLKRFEQOQSOQPOFQEQCHQHOE	135
QY	54	-----REDKRYE-----RKKEED 68	
DB	136	QRPERRKQOCVRCRCRYRQDENPWRREER 164	

```
RESULT 2
VCLA_GOSHI STANDARD: PRT: 605 AA.
AC P09799:
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE VICILIN GC72-A PRECURSOR (ALPHA-GLOBULIN A).
OS Gossypium hirsutum (Upland cotton).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Malvales; Malvaceae; Gossypium.
RN [1]
RP SEQUENCE FROM N.A.
RA Chlan C.A., Borroto K., Kamalay J.A., Dure L. III,
RT "Developmental biochemistry of cottonseed embryogenesis and
RT germination. XIX. Sequences and genomic organization of the alpha
RT globulin (vicilin) genes of cottonseed."
RL Plant Mol. Biol. 9:533-546(1987).
CC -1- FUNCTION: SEED STORAGE PROTEIN.
CC -1- SUBCELLULAR LOCATION: COTYLEDONARY MEMBRANE-BOUND VACUOLAR PROTEIN
CC BODIES.
CC -1- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC CONVICILIN, CONGLYCININ, ETC.).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M19378; AAA33069.1; -
DR PIR: S06398; S06398.
DR HSSP: P50477; ICAX.
DR INTERPRO: IPR001113; -
DR PFAM: PF00546; Seedstore_7s; 1.
DR KW Seed storage protein; Signal.
FT SIGNAL 1 23
FT CHAIN 24 605 VICILIN GC72-A
FT SEQUENCE 605 AA; 71049 MW; C9DB9371C976953B CRC64;
SO
OY 2 RORPQOQYEQCQRCQNHETPRHMQTCQRCRRYE---KEKRRKQKRYEQQREDE 57
Db 114 QQQQKQFKKQECQRCQNHETPRHMQTCQRCRRYE---KEKRRKQKRYEQQREDE 173
OY 58 -----EKYEPMKEE 67
Db 174 GEGQQRNNPYTFHRSFQERFREE 197
RESULT 3
TRHY_HUMAN STANDARD: PRT: 1898 AA.
AC Q07283:
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE TRICHOHYALIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93280194; PubMed=7685034;
```

```
RA Lee S.-C., Kim I.-G., Marekov L.N., O'Keefe E.J., Parry D.A.D.,
RA Steinert P.M.:
RT "The structure of human trichohyalin. Potential multiple roles as a
RT functional EF-hand-like calcium-binding protein, a cornified cell
RT envelope precursor, and an intermediate filament-associated (cross-
RT linking) protein."
RL J. Biol. Chem. 268:12164-12176(1993).
RN [2]
RP SEQUENCE OF 1731-1898 FROM N.A., AND CHARACTERIZATION.
RX MEDLINE=93315897; PubMed=7686953;
RA O'Keefe E.J., Hamilton E.H., Lee S.-C., Steinert P.M.:
RT "Trichohyalin: a structural protein of hair, tongue, nail, and
RT epidermis."
RL J. Invest. Dermatol. 101:655-715(1993).
CC -1- FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES
CC IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE
CC INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR
CC LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY
CC ISODIPEPTIDE BONDS. IT MAY SERVE AS SCAFFOLD PROTEIN, TOGETHER
CC WITH INVOLUCRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN
CC ANCHOR THE CELL ENVELOPE TO THE KIF NETWORK. IT MAY BE INVOLVED IN
CC ITS OWN CALCIUM-DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL
CC DIFFERENTIATION.
CC -1- SUBUNIT: MONOMER (PROBABLE).
CC -1- TISSUE SPECIFICITY: FOUND IN THE HARD KERATINIZING TISSUES SUCH AS
CC THE INNER ROOT SHEATH (IRS) OF HAIR FOLLICLES AND MEDULLA, AND IN
CC THE FILIFORM PAPILLAE OF DORSAL TONGUE EPITHELIUM (PROBABLE).
CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING LATE DIFFERENTIATION OF
CC THE EPIDERMIS.
CC -1- DOMAIN: CONSISTS OF NINE DOMAINS. DOMAIN 1 CONTAINS TWO EF-HAND
CC CALCIUM-BINDING DOMAINS. DOMAINS 2-4, 6, AND 8 ARE ALMOST
CC ENTIRELY ALPHA-HELICAL, CONFIGURED AS A SERIES OF PEPTIDE REPEATS
CC OF VARYING REGULARITY, AND ARE THOUGHT TO FORM A SINGLE-STANDED
CC ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS
CC THE MOST REGULAR AND MAY BIND KIF DIRECTLY BY IONIC INTERACTIONS.
CC DOMAINS 5 AND 7 ARE LESS WELL ORGANIZED AND MAY INDUCE FOLDS IN
CC THE MOLECULE. DOMAIN 9 CONTAINS THE C-TERMINUS, CONSERVED AMONG
CC DIFFERENT SPECIES.
CC -1- PTM: KNOWN SUBSTRATE OF TRANSGLUTAMINASE. SOME 200 ARGININES ARE
CC PROBABLY CONVERTED TO CITRULLINES BY PEPTIDYLARGININE DEIMINASE.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE S-100
CC FAMILY.
CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L09190; AAA65582.1; -
DR PIR: A45973; A45973.
DR HSSP: P02633; IBOC.
DR MIM: 190370; -
DR INTERPRO: IPR001751; -
DR INTERPRO: IPR002048; -
DR PFAM: PF01023; S.100; 1.
DR PFAM: PF00036; efhand; 1.
DR PROSITE: PS00018; EF_HAND; 1.
DR PROSITE: PS00303; S100_CABP; 1.
KW Repeat; Calcium-binding.
FT DOMAIN 1 91
FT CA_BIND 22 33
FT CA_BIND 62 73
FT DOMAIN 314 390
FT REPEAT 314 326
FT REPEAT 327 339
FT REPEAT 340 351
FT REPEAT 352 364
FT REPEAT 365 377
```

```

FT REPEAT 378 390 1-6.
FT DOMAIN 391 444 9 X 6 AA TANDEM REPEATS OF R-R-E-Q-Q-L.
FT REPEAT 396 402 2-1.
FT REPEAT 397 402 2-2.
FT REPEAT 403 408 2-3.
FT REPEAT 409 414 2-4.
FT REPEAT 415 420 2-5.
FT REPEAT 421 426 2-6.
FT REPEAT 427 432 2-7.
FT REPEAT 433 438 2-8.
FT REPEAT 439 444 2-9.
FT DOMAIN 444 702 9 X 28 AA APPROXIMATE TANDEM REPEATS.
FT DOMAIN 923 1162 8 X 30 AA TANDEM REPEATS.
FT REPEAT 923 952 4-1.
FT REPEAT 923 952 4-2.
FT REPEAT 953 982 4-3.
FT REPEAT 983 1012 4-4.
FT REPEAT 1013 1042 4-5.
FT REPEAT 1043 1072 4-6.
FT REPEAT 1073 1102 4-7.
FT REPEAT 1103 1132 4-8.
FT DOMAIN 1133 1162 23 X 26 AA APPROXIMATE TANDEM REPEATS.
FT DOMAIN 1250 1849 F -> L (IN REF. 2).
FT CONFLICT 1752 1752 Q -> K (IN REF. 2).
FT CONFLICT 1794 1801 V -> G (IN REF. 2).
FT CONFLICT 1857 1880
FT CONFLICT 1880
SQ SEQUENCE 1898 AA; 247219 MW; A74B5947FB62E31D CRC64;

```

```

Query Match 29.4%; Score 113; DB 1; Length 1898;
Best Local Similarity 33.3%; Pred. No. 0.03;
Matches 25; Conservative 22; Mismatches 20; Indels 8; Gaps 2;

```

```

QY 2 RQDPQOQYECOCRCORHTEPRHMQTCOQRCERYEK-----EKKROOKRYEQ-Q 53
Db 468 RQDEERDMKRELETERHQRQKQKLRQDEERERKMKLEERERQOQRRQOLR 527
QY 54 RQDEKRYEERMKRED 68
Db 528 RQDEERERQRLKROE 542

```

```

RESULT 4
TRHY_RABIT STANDARD: PRT: 1407 AA.
AC P37709;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE TRICHOHYALIN.
GN THH.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN (1)
RP SEQUENCE FROM N.A.
RA Fietz M.J., Rogers G.E.;
RL Submitted (DEC-1992) to the EMBL/Genbank/DBJ databases.
CC -I- FUNCTION: INTERMEDIATE FILAMENT-ASSOCIATED PROTEIN THAT ASSOCIATES
CC IN REGULAR ARRAYS WITH KERATIN INTERMEDIATE FILAMENTS (KIF) OF THE
CC INNER ROOT SHEATH CELLS OF THE HAIR FOLLICLE AND THE GRANULAR
CC LAYER OF THE EPIDERMIS. IT LATER BECOMES CROSS-LINKED TO KIF BY
CC ISODIPEPTIDE BONDS. IT MAY SERVE AS SCAFFOLD PROTEIN, TOGETHER
CC WITH INVOLUCRIN, IN THE ORGANIZATION OF THE CELL ENVELOPE OR EVEN
CC ANCHOR THE CELL ENVELOPE TO THE KIF NETWORK. IT MAY BE INVOLVED IN
CC ITS OWN CALCIUM-DEPENDENT POSTSYNTHETIC PROCESSING DURING TERMINAL
CC DIFFERENTIATION.
CC -I- SUBUNIT: HOMODIMER (PROBABLE).
CC -I- TISSUE SPECIFICITY: FOUND IN THE HARD KERATINIZING TISSUES SUCH AS
CC THE INNER ROOT SHEAT (IRS) OF HAIR FOLLICLES AND MEDULLA, AND IN
CC THE FILIFORM PAPILLAE OF DORSAL TONGUE EPITHELIUM (PROBABLE).
CC -I- DEVELOPMENTAL STAGE: EXPRESSED DURING LATE DIFFERENTIATION OF
CC THE EPIDERMIS.

```

```

CC -I- DOMAIN: CONSISTS OF NINE DOMAINS. DOMAIN 1 CONTAINS TWO EF-HAND
CC CALCIUM-BINDING DOMAINS. DOMAINS 2-4, 6, AND 8 ARE ALMOST
CC ENTIRELY ALPHA-HELICAL, CONFIGURED AS A SERIES OF PEPTIDE REPEATS
CC OF VARYING REGULARITY, AND ARE THOUGHT TO FORM A SINGLE-STRANDED
CC ALPHA-HELICAL ROD STABILIZED BY IONIC INTERACTIONS. DOMAIN 6 IS
CC THE MOST REGULAR AND MAY BIND KIF DIRECTLY BY IONIC INTERACTIONS.
CC DOMAINS 5 AND 7 ARE LESS WELL ORGANIZED AND MAY INDUCE FOLDS IN
CC THE MOLECULE. DOMAIN 9 CONTAINS THE C-TERMINUS, CONSERVED AMONG
CC DIFFERENT SPECIES.
CC -I- PTM: KNOWN SUBSTRATE OF TRANSGLUTAMINASE. SOME 200 ARGININES ARE
CC PROBABLY CONVERTED TO CITRULLINES BY PEPTIDYLARGININE DEIMINASE.
CC -I- SIMILARITY: IN THE N-TERMINAL SECTION, BELONGS TO THE S-100
CC FAMILY.
CC -I- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z19092; CAA79519.1; -.
CC PIR: S28589; S28589.
CC HSSP: P02633; 1B0C.
CC DR INTERPRO: IPR001751; -.
CC DR INTERPRO: IPR002048; -.
CC DR PFAM: PF01023; S_100; 1.
CC DR PFAM: PF00036; ethand; 1.
CC DR PROSITE: PS00018; EF_HAND; 1.
CC DR PROSITE: PS00303; S100_CABP; 1.
CC KW Repeat: Calcium-binding.
CC FT DOMAIN 1 91 S-100 LIKE.
CC FT CA_BIND 22 33 SITE I (LOW AFFINITY) (POTENTIAL).
CC FT CA_BIND 62 73 SITE II (HIGH AFFINITY) (POTENTIAL).
CC SQ SEQUENCE 1407 AA; 183781 MW; AE17D2A159F12B7F CRC64;

```

```

Query Match 28.6%; Score 110; DB 1; Length 1407;
Best Local Similarity 36.5%; Pred. No. 0.04;
Matches 27; Conservative 19; Mismatches 18; Indels 10; Gaps 2;

```

```

QY 3 QRDPOQYECOCRCORHTEPRHMQTCOQRCERYEKRRQOK-----RYEQR 54
Db 256 QQLRRELEIRERERQDEERREO-QLRERQRLQDEERERQQLRLEIRERERQ 313
QY 55 EDEKRYEERMKRED 68
Db 314 RQDEERERQRLQOE 327

```

```

RESULT 5
UR2_HUMAN STANDARD: PRT: 482 AA.
AC Q15696;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE U2 SMALL NUCLEAR RIBONUCLEOPROTEIN AUXILIARY FACTOR 35 KDA SUBUNIT
DE RELATED-PROTEIN 2.
GN U2AF1-RS2 OR U2AF1RS2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=BRIN;
RX MEDLINE=96163878; PubMed=8586425;
RA Kitagawa K., Wang X., Hatada T., Yamoka T., Nojima H.,
RA Inazawa J., Abe T., Mitsuya K., Oshimura M., Murata A., Monden M.,
RA Mukai T.;
RT *Isolation and mapping of human homologues of an imprinted mouse gene

```


CC MATURE PHLOEM AND THE CELLS OF DEVELOPING COTYLEDONS.
 CC - DEVELOPMENTAL STAGE. IN THE COTYLEDON, EXPRESSION IS NOT DETECTED
 CC UNTIL 10 DAYS AFTER FERTILIZATION. BETWEEN 10-19 DAYS AFTER
 CC FERTILIZATION, EXPRESSION INCREASES RAPIDLY BUT DECLINES 20-30
 CC DAYS AFTER FERTILIZATION. 30 DAYS AFTER FERTILIZATION, NO
 CC EXPRESSION OCCURS. THIS EXPRESSION PATTERN CLOSELY PARALLELS THE
 CC RATE OF SUCROSE UPTAKE IN THE COTYLEDON.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L06038; AAB03894.1; -
 CC PIR: J01730; J01730.
 CC DR HSSP: P50477; ICAM.
 CC DR INTERPRO: IPR001113; -
 CC DR PFAM: PF00546; Seedstore_7s; 1.
 CC KW Transport: Sugar transport; Signal; Membrane.
 CC FT SIGNAL 1 29
 CC FT CHAIN 30 524 SUCROSE-BINDING PROTEIN.
 CC FT SEQUENCE 524 AA: 60522 MW: 0251EP90796EF341 CRC64;
 SQ

Query Match 25.0%; Score 96; DB 1; Length 524;
 Best Local Similarity 35.9%; Pred. No. 0.23;
 Matches 23; Conservative 16; Mismatches 15; Indels 10; Gaps 3;

QY 12 OCQERQCHETPRHMQTCQRCER-RYEKKERKQ-----QKRYEQDQDEDEKYEERM 64
 DB 46 OCOQOQOQTEGCDK-R--VCLQSCDRYHMKQREKQIQETREKKKEEEREREQOQCH 102
 QY 65 KED 68
 DB 103 EED 106

RESULT 10
 ID CK11_YEAST STANDARD: PRT: 538 AA.
 AC P23291;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE CASEIN KINASE I HOMOLOG 1 (EC 2.7.1.1-).
 GN YCK1 OR CK12 OR YHR135C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 OC Saccharomycetaceae; Saccharomyces.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C;
 RX MEDLINE=92108037; PubMed=1729698;
 RA Robinson L.C., Hubbard E.D.A., Graves P.R., de Paoli-Roach A.A.,
 RA Roach P.J., Kung C., Haas D.W., Hagedorn C.H., Goebel M.,
 RA Culbertson M.R., Carlson M.;
 RT "Yeast casein kinase I homologues: an essential gene pair."
 RL Proc. Natl. Acad. Sci. U.S.A. 89:28-32(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92329955; PubMed=1627830;
 RA Wang P.-C., Vancura A., Mitcheson T.G.M., Kuret J.;
 RT "Two genes in Saccharomyces cerevisiae encode a membrane-bound form
 of casein kinase-1."
 RL Mol. Biol. Cell 3:275-286(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C / AB972;
 RX MEDLINE=94378003; PubMed=8091229;
 RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,

RA Du Z., Favella A., Fulton L., Gattung S., Giesel C., Kirsten J.,
 RA Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
 RA Lacroix P., Louis E.J., Macri C., Maris E., Meneses S., Mouser L.,
 RA Nhan M., Rifkin L., Riles L., St Peter H., Treaskis E., Vaughan K.,
 RA Vignati D., Wilcox L., Wohldman P., Waterston R., Wilson R.,
 RA Vaudin M.;
 RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
 RT VIII."
 RL Science 265:2077-2082(1994).
 RN [4]
 RP SEQUENCE OF 1-38 FROM N.A.
 RC STRAIN-SIGMA 1278B;
 RX MEDLINE=98151783; PubMed=9491083;
 RA Iragui I., Vissers S., Carlioux M., Urrestarazu A.;
 RT Characterisation of Saccharomyces cerevisiae ARO8 and ARO9 genes
 RT encoding aromatic aminotransferases I and II reveals a new
 RT aminotransferase subfamily."
 RL Mol. Gen. Genet. 257:238-248(1998).
 CC - FUNCTION: CASEIN KINASES ARE OPERATIONALLY DERIVED BY THEIR
 CC PREFERENTIAL UTILIZATION OF ACIDIC PROTEINS SUCH AS CASEINS
 CC AS SUBSTRATES.
 CC - SUBCELLULAR LOCATION: PLASMA-MEMBRANE BOUND.
 CC - SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC CASEIN KINASE I SUBFAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M74552; AAA35229.1; -
 CC EMBL: X60327; CAA42897.1; -
 CC EMBL: U10398; AAB68417.1; -
 CC EMBL: Y13625; CAA73948.1; -
 CC PIR: B43764; B43764.
 CC PIR: S29521; S29521.
 CC PIR: S48979; S48979.
 CC HSSP: P40233; 2CSN.
 CC SGP: S0001177; YCK1.
 CC INTERPRO: IPR000719; -
 CC INTERPRO: IPR002290; -
 CC PFAM: PF00069; Pkinase; 1.
 CC PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 CC PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
 CC PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 CC KW Transferase; Serine/threonine-protein kinase; ATP-binding;
 CC Multigene family; Membrane; Prenylation; Lipoprotein.
 CC DOMAIN 69 353
 CC NP_BIND 75 83
 CC BINDING 98 98
 CC ACT_SITE 188 188
 CC LIPID 537 537
 CC LIPID 538 538
 CC CONFLICT 367 367
 CC SEQUENCE 538 AA: 61715 MW: 77344717818C2D17 CRC64;

Query Match 24.9%; Score 95.5; DB 1; Length 538;
 Best Local Similarity 26.0%; Pred. No. 0.26;
 Matches 19; Conservative 26; Mismatches 21; Indels 7; Gaps 2;

QY 2 RQRDQOQOYEQOCQRCQRR-----ETPRHMQTCQRCERRY--EKRRKQOQRYEEOOR 54
 DB 403 QOQOQOQOYQAQTEQDMNSQYKRYDPTSYEAUHQOQKLEQOQRKQOQOQKLEQOL 462
 QY 55 EDEKRYEERKKEE 67
 DB 463 QEOQLQOQLOQOQOQOQ 475

```

RESULT 11
GARP_PLAAF
ID GARP_PLAAF STANDARD: PRT: 678 AA.
AC P13816;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GLUTAMIC ACID-RICH PROTEIN PRECURSOR.
GN GARP.
OS Plasmodium falciparum (isolate FC27 / Papua New Guinea).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
RN [1]
RP MEDLINE-89040046; PubMed-2903445;
RX Trigila T., Stahl H.-D., Crewther P.E., Silva A., Anders R.F.,
RA Kemp D.J.:
RT "Structure of a Plasmodium falciparum gene that encodes a glutamic
acid-rich protein (GARP)".
RL Mol. Biochem. Parasitol. 31:199-202(1988).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL collaboration -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J03998; AAA29605.1; -.
DR PIR; A54514; A54514.
KW Repeat; Malaria; Antigen; Signal.
FT SIGNAL 1 25
FT CHAIN 26 678 GLUTAMIC ACID-RICH PROTEIN.
FT DOMAIN 120 164 15 X 3 AA TANDEN REPEATS OF K-K-X.
FT DOMAIN 372 416 9 X APPROXIMATE TANDEN REPEATS.
FT DOMAIN 417 441 5 X APPROXIMATE TANDEN REPEATS.
FT DOMAIN 576 604 POLY-GLU.
FT DOMAIN 605 653 7 X APPROXIMATE TANDEN REPEATS.
FT DOMAIN 654 663 POLY-GLU.
SQ SEQUENCE 678 AA; 80551 MW; 2A8F85606496E93E CRC64;

Query Match 24.6%; Score 94.5; DB 1; Length 678;
Best Local Similarity 26.5%; Pred. NO. 0.37;
Matches 18; Conservative 23; Mismatches 20; Indels 7; Gaps 1;

OY 1 NROSDPOQVYDCCORCRHFTFRPHMOTCCQRCRRYRKEKRKQKRYEPOQRDEEY 60
DB 546 DKEEDKKEESKEVDF-----ESKEVDEDEVEDEDEDEDEDEDEDEDEDEDE 598
QY 61 EERKKEED 68
DB 599 EEEEDDED 606

RESULT 12
INCE_CHICK
ID INCE_CHICK STANDARD: PRT: 877 AA.
AC P53352;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE INNER CENTROMERE PROTEIN (INCENP).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
RN [1]
RP MEDLINE-94012983; PubMed-8408220;
RX Mackay A.M., Eckley D.M., Chue C., Earnshaw W.C.:
RT "Molecular analysis of the INCENPs (inner centromere proteins):
separate domains are required for association with microtubules

```

[illegible]

CC	BHLH PROTEIN.
CC	-1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC	-1- TISSUE SPECIFICITY: UNBOUTHOUSLY EXPRESSED IN THE EMBRYO.
CC	-1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF TRANSCRIPTION FACTORS. HIGHEST TO HLF-1 ALPHA.
CC	-1- SIMILARITY: CONTAINS 1 PAS (PEER-ARN-TSM) DIMERIZATION DOMAIN.
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC between the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC	EMBL: U43090; AAC47303.1; -.
DR	FLYBASE: FBgn0015542; sima.
DR	INTERPRO: IPR000014; -.
DR	INTERPRO: IPR001610; -.
DR	INTERPRO: IPR003015; -.
DR	PFAM: PF00785; PAC; 1.
DR	PFAM: PF00989; PAS; 2.
DR	PROSITE: PS00038; HELIX LOOP HELIX; 1.
KW	Repeat; DNA-binding; Nuclear protein; Transcription regulation; Activator; Coiled coil.
FT	DNA_BIND 72 85 BASIC DOMAIN.
FT	DOMAIN 86 126 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT	REPEAT 170 236 PAS-1.
FT	REPEAT 309 375 PAS-2.
FT	DOMAIN 381 422 PAC MOTIF.
FT	DOMAIN 577 587 PRO-RICH.
FT	DOMAIN 883 904 COILED COIL (POTENTIAL).
FT	DOMAIN 987 1058 COILED COIL (POTENTIAL).
FT	DOMAIN 1114 1160 COILED COIL (POTENTIAL).
FT	DOMAIN 26 37 POLY-SER.
FT	DOMAIN 718 725 POLY-SER.
FT	DOMAIN 759 763 POLY-SER.
FT	DOMAIN 767 779 POLY-SER.
FT	DOMAIN 889 899 POLY-SER.
FT	DOMAIN 911 922 POLY-SER.
FT	DOMAIN 949 952 POLY-SER.
FT	DOMAIN 994 1002 POLY-SER.
FT	DOMAIN 1024 1042 POLY-SER.
FT	DOMAIN 1117 1128 POLY-SER.
FT	DOMAIN 1148 1160 POLY-SER.
FT	DOMAIN 1203 1206 POLY-SER.
FT	DOMAIN 1275 1282 POLY-SER.
FT	DOMAIN 1296 1299 POLY-SER.
FO	SEQUENCE 1505 AA: 165686 MW: 165686 NM: ECAPD00E7A49B29A CRC64: 1505 AA: 165686 NM: ECAPD00E7A49B29A CRC64:

```

Query Match          24.6%; Score 94.5; DB 1, Length 1505;
Best Local Similarity 28.8%; Pred. No. 0.74;
Matches      21; Conservative    18; Mismatches   15; Indels   19; Gaps    2.

QY       11 EOCERCORH-----TEPRHMOTCOQRCEERYEKERKROQRYEBOQREDE----- 57
           ||| ||| |||
Db        987 EECHNRLLQQHQDQDQQTSGNERFTTQQLLQDELQLQLEBQDQRRQDQDQDQDQDQDQDQLLST 1046

QY       58 -----EKYEERM 64
           ||| : |
Db        1047 NIECKKEKYDVQM 1059

RESULT 14
GLT_DROME
ID      GLT_DROME      STANDARD;      PRT; 1023 AA.
AC      P33438;
DT      01-FEB-1994 (Rel. 28, Created)
DT      01-FEB-1994 (Rel. 28, Last sequence update)
DT      30-MAY-2000 (Rel. 39, Last annotation update)
DE      GLUTACTIN PRECURSOR.
LN      GLUT.

```

OS	Drosophila melanogaster (Fruit fly).
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC	Ephyroidae; Drosophilidae; Drosophila.
RN	(1)
RP	SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RC	STRAIN-OREGON-R;
RX	MEDLINE=90214632; PubMed=2108864;
RA	Olson P.F., Fessler L.I., Nelson R.E., Sterne R.E., Campbell A.G.,
RA	Fessler J.H.,
RT	"Glutactin", a novel Drosophila basement membrane-related glycoprotein
RT	with sequence similarity to serine esterases";
RL	EMBO J. 9:1219-1227(1990).
CC	- FUNCTION: NOT KNOWN. BINDS CALCIUM IONS.
CC	- SUBCELLULAR LOCATION: BASAMENT MEMBRANES.
CC	- 1-PMM: EXTENSIVELY O-GLYCOSYLATED AND ALSO N-GLYCOSYLATED.
CC	- 1-PMM: FOUR TYROSINES ARE SULFATED.
CC	- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE TYPE-B
CC	CARBOXYLESTERASE/LIPASE FAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL: X53286; CAA37380.1; -.
DR	HSSP: P21836; IMAH.
DR	FLYBASE: FBgn0001114; Glt.
DR	INTERPRO: IPR002018; -.
DR	PFAM: PF00135; Coesterase; 2.
DR	PROSITE: PS00941; CARBOXYLESTERASE_B_2; 1.
KW	Glycoprotein; Sulfatation; Calcium-Binding; Signal.
FT	SIGNAL 1 17
FT	CHAIN 18 1023 GLUTACTIN.
FT	SIMILAR 18 602 WITH TYPE-B CARBOXYLESTERASE/LIPASES.
FT	DOMAIN 603 615 POLY-THR.
FT	DOMAIN 616 1023 GLU/GLN-RICH.
FT	CARBOHD 115 115 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHD 368 368 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHD 402 402 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHD 810 810 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	DISULFD 123 145 BY SIMILARITY.
FT	DISULFD 298 316 BY SIMILARITY.
SO	SEQUENCE 1023 AA; 118412 MW; 3638CF9A5860E8C CRC64;

Query Match	24.5%	Score	94	DB 1	Length	1023
Best Local Similarity	37.3%	Pred. No.	0.56			
Matches	28	Conservative	16	Mismatches	23	Indels
					8	Gaps
					4	

QY	1	NKRDPPQOO-YEOCQER-ORHETEPRIHQCOQ----	RCERYEKEKROOKR-YEEQ	52
	::: :			::: :
Db	671	DQGRPDQEPDQDERLRQDQDEERLRQDQDEERLRQDRELEERLRQDQDERQYERE		730

QY	53	QREDEKTEERAKKE	67
	::: :		::: :
Db	731	QOERQERERELEERQ	745

RESULT	15
YAFB_SCHPO	
ID	YAFB_SCHPO
AC	Q09863
DT	01-FEB-1996 (Rel. 33, Created)
DT	01-FEB-1996 (Rel. 33, Last sequence update)
DT	01-OCT-1996 (Rel. 34, Last annotation update)
DE	HYPOTHEETICAL 122.9 KDA PROTEIN C29B6.10C IN CHROMOSOME I.
GN	SPAC29B6.10C.
OS	Schizosaccharomyces pombe (fission yeast).
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;

OC Schizosaccharomycetaceae; Schizosaccharomyces.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RA Jones L., Murphy L., McNeil A., Simpson I., Harris D., Barrell B.G.,
 RA Rajandream M.A., Walsh S.V.;
 RL Submitted (Oct-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: TO YEAST YNL091W.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z66525; CAA91432.1; -.
 KW Hypothetical protein.
 FT DOMAIN 33 46 POLY-LYS.
 FT DOMAIN 184 194 POLY-LYS.
 FT DOMAIN 695 698 POLY-GLU.
 FT DOMAIN 714 717 POLY-GLU.
 FT DOMAIN 718 721 POLY-ARG.
 FT DOMAIN 938 941 POLY-SER.
 SO SEQUENCE 1085 AA: 122940 MW: 041164132676F233 CRC64;

Query Match 24.5%; Score 94; DB 1; Length 1085;
 Best Local Similarity 29.1%; Pred. No. 0.61; 21; Indels 14; Gaps 2;
 Matches 23; Conservative 21; Mismatches 2;

OY 2 RORDPOOQYEOC-OERCQRIETE-----PRIMQTCQRCRCRRYKEREKROOK 47
 DB 596 KRRDKKKQKLAKKEERQRRERELAQAAQKALEAKKROFEARKKREQRLKRRQEKQ 655
 OY 48 RYEEQOREDEPEKYEERKE 66
 DB 656 ELEROKREEROKOKEREKK 674

Search completed: March 1, 2001, 16:16:41
 Job time: 421 sec

